Kybernetika

Lubomír Ohera

On the problem of separability of some pattern recognition learning systems

Kybernetika, Vol. 5 (1969), No. 5, (420)--436

Persistent URL: http://dml.cz/dmlcz/125471

KYBERNETIKA ČÍSLO 5, ROČNÍK 5/1969

On the Problem of Separability of Some Pattern Recognition Learning Systems

LUBOMÍR OHERA

The separating capacity of a certain class of pattern recognition learning systems is investi gated. The systems of that kind learn from given labeled sample patterns in such a way that afte the learning period is over all the given sample patterns are placed exactly to the same category as they were labeled to belong to. This paper gives estimations of the upper bound of the number of dichotomies separable by those systems.

Many pattern recognition learning systems are based on the following principle. The learning system is given labeled sample patterns from all categories and then during the learning period the system is to find the proper values of variable parameters (often called weights) so that after the learning period is over the system places all the given sample patterns exactly to the same category as they were labeled to belong to.

This very broad class of pattern recognition learning systems includes such well-known examples as simple perceptrons, perceptrons with some weights of connections randomly chosen before the beginning of the learning period, layered machines, ADALINE and MADALINE (as far as the exact separation of all categories is required after the learning period), etc.

In order to learn general regularities of this class of pattern recognition learning systems, there are, besides practical realization, two main problems to be investigated:

1. to decide whether with a given structure of the learning system all possible situations may be learned, i.e. whether for any of all possible configurations of labeled sample patterns there exists a set of values of variable parameters such that the system separates the sample patterns in such a way that the decision of the system is always strictly in accordance with the labeling of the patterns;

2. to study the learning period, i.e. to find an algorithm by means of which the system with the given structure may be taught the situation provided that the given situation is separable by the structure of the given system.

The second problem was studied by many researchers [1–3] and many interesting and important results have been obtained to this date. Leaving that problem apart, this paper deals with the first problem.

Let us have two sets of vectors in a subspace χ of an n-dimensional Euclidean space and let us denote

$$^1\mathbf{Z}_1, \ldots, {}^1\mathbf{Z}_{p_1}$$

those of them which belong to category w_1, and

$$^2\mathbf{Z}_1, \ldots, {}^2\mathbf{Z}_{p_2}$$

those which belong to category w_2. Let us further define discriminant function G as follows.

Definition 1. We shall call discriminant function G any scalar-valued function defined over the entire subspace χ and satisfying the following system of inequalities:

$$G(^1\mathbf{Z}_i) > 0\,, \quad i = 1, \ldots, p_1\,, \tag{1}$$
$$G(^2\mathbf{Z}_i) < 0\,, \quad i = 1, \ldots, p_2\,.$$

We shall now define the decision according to the discriminant function in the following way. Having chosen a function G satisfying (1), we may compute for any given vector $\mathbf{X}$ from χ the value of $G(\mathbf{X})$ and decide as follows:

$$\text{if } G(\mathbf{X}) > 0 \quad \text{then } \mathbf{X} \text{ belongs to } w_1$$

and

$$\text{if } G(\mathbf{X}) < 0 \quad \text{then } \mathbf{X} \text{ belongs to } w_2\,.$$

Note. Equation $G(\mathbf{X}) = 0$ defines a boundary between all vectors $\mathbf{X}$ which are decided upon as belonging to category w_1 and all vectors $\mathbf{X}$ which are decided upon as belonging to category w_2. As for the points lying on the boundary, we may either reject them, i.e. let them undecided upon, or we may place them into any of the two categories. If $\mathbf{X}$ is one of sample pattern vectors, the decision is always correct, as follows directly from (1), and in case of the other vectors of χ the decision depends on the choice of function G, i.e. on the choice of the decision boundary.

We shall now generalize the considerations for k categories. Let us denote the vectors belonging to category w_j

$$^j\mathbf{Z}_1, \ldots, {}^j\mathbf{Z}_{p_j}\,, \quad j = 1, \ldots, k\,.$$

We shall call a set of discriminant functions any set of scalar-valued functions $G_1, \ldots, G_\nu$ defined over the entire subspace χ, where any of the discriminant functions satisfies an appropriate system of inequalities. In order to find functions G_h, we divide all categories into three groups A_h, B_h and C_h. Not considering group C_h, we find a boundary between groups A_h and B_h in the very same manner as if we had only two categories. Discriminant function G_h is therefore any scalar-valued function satisfying the following system of inequalities:

(2) $$G_h({}^{A_h}\boldsymbol{Z}) > 0\,,$$

(3) $$G_h({}^{B_h}\boldsymbol{Z}) < 0\,,$$

where (2) must be satisfied for all vectors of all categories belonging to group A_h, and (3) for all vectors of all categories belonging to group B_h.

Having done this procedure for $h = 1, \ldots, v$, we can write the truth table for categories $w_1, \ldots, w_k$ and for discriminant functions $G_1, \ldots, G_v$ (Table 1). The columns

Table 1.

	w_1	w_2	$\ldots$	w_k
G_1	${}^1b_1$	${}^1b_2$	$\ldots$	${}^1b_k$
G_2	${}^2b_1$	${}^2b_2$	$\ldots$	${}^2b_k$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
G_v	${}^vb_1$	${}^vb_2$	$\ldots$	${}^vb_k$

of the truth table are the code words of categories corresponding to our system of dividing categories into groups; ${}^xb_y$ are binary numbers "0" or "1", or "d", where d's mean don't care conditions introduced in switching theory. Numbers 1, 0 and d stand for the fact that category w_y belongs to group A_x, B_x and C_x, respectively.

The code chosen for the description of categories must not be ambiguous, i.e. any combination of v binary numbers 1 and 0 must indicate at most one category. Since the number of possible combinations of v binary numbers may be greater than the number of code words indicating categories (taking into consideration that a code word containing "d" is an abbreviation for two code words indicating the same category, one of them having "1" and the other "0" in the same position, where "d" occurred), we can, given a combination of v binary numbers obtained as the result of the investigation of a given vector $\boldsymbol{X}$ be means of functions $G_1, \ldots, G_v$, either decide which category this v-dimensional binary vector corresponds to, and thus which category the given vector $\boldsymbol{X}$ belongs to, or finding no category with the same code, let the vector $\boldsymbol{X}$ be undecided upon.

Note. Instead of the set of discriminant functions, another set, the set of functions $H_1, \ldots, H_k$ may be introduced, where the functions $H_1, \ldots, H_k$ are defined over the entire subspace χ and satisfy the following system of inequalities:

$$H_r({}^r\boldsymbol{Z}_i) > H_s({}^r\boldsymbol{Z}_i)\,, \quad r = 1, \ldots, k, \; s = 1, \ldots, k, \; s \neq r, \; i = 1, \ldots, p_r\,.$$

Choosing a convenient system of dividing the categories into groups A_h, B_h and C_h, the system of discriminant functions can be readily obtained.

In accordance with various ways of dividing categories into groups we have various code words corresponding to categories $w_1, \ldots, w_k$. We shall mention only three of them. Table 2 shows the truth table for storing boundaries between any pair of categories, while Table 3 shows the truth table for storing for every category

Table 2.

	w_1	w_2	w_3	$\ldots$	w_{k-2}	w_{k-1}	w_k
G_1	1	0	d	$\ldots$	d	d	d
G_2	1	d	0	$\ldots$	d	d	d
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$G_{\nu-1}$	d	d	d	$\ldots$	1	d	0
G_ν	d	d	d	$\ldots$	d	1	0

Table 3.

	w_1	w_2	w_3	$\ldots$	w_k
G_1	1	0	0	$\ldots$	0
G_2	0	1	0	$\ldots$	0
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
G_k	0	0	0	$\ldots$	1

Table 4.

	w_1	w_2	w_3	w_4	w_5	w_6	w_7
G_1	0	0	0	1	1	1	1
G_2	0	1	1	0	0	1	1
G_3	1	0	1	0	1	0	1

the boundary between this category and all remaining categories. If we want to decrease the number of boundaries as much as possible, we use one of the codes allowing encoding k categories by means of ε code words, where

$$\varepsilon - 1 < \log_2 k \leqq \varepsilon .$$

One of the possibilities to accomplish it is to prescribe to every category the code word that can be read as the name of the category in binary numbers. For 7 categories, for example, the code words are shown in Table 4.

In order to solve our system of inequalities, we have to admit only special families of functions so that the boundaries will be, for example, hyperplanes, hyperspheres, piecewise linear hypersurfaces or so.

Before treating separating capacity of some pattern recognition systems of this kind we shall define some important terms and prove a few theorems.

Definition 2. Let us have two categories w_1 and w_2 defined as follows:

$$w_1 \equiv \{{}^1\mathbf{Z}_i\,, \quad i = 1, \ldots, p_1\}\,, \tag{4}$$

$$w_2 \equiv \{{}^2\mathbf{Z}_i\,, \quad i = 1, \ldots, p_2\}\,. \tag{5}$$

The two categories are said to be homogeneously linearly separable if there exists a vector $\mathbf{g}$ satisfying

$$\mathbf{g}\,.\,{}^1\mathbf{Z}_i > 0\,, \quad i = 1, \ldots, p_1\,, \tag{6}$$

$$\mathbf{g}\,.\,{}^2\mathbf{Z}_i < 0\,, \quad i = 1, \ldots, p_2\,. \tag{7}$$

Note. Vector $\mathbf{g}$ defines separating hyperplane

$$\mathbf{g}\,.\,\mathbf{X} = 0\,, \tag{8}$$

which passes through the origin of the space and is perpendicular to $\mathbf{g}$, which is clearly seen from (8).

Having p points in χ, there are 2^p possibilities how to divide the points into two categories. Any of these possibilities will be called a dichotomy. If the points belonging to categories w_1 and w_2 are defined by (4) and (5), $(p = p_1 + p_2)$, we denote this dichotomy as

$$\left[\bigcup_{i=1}^{p_1} {}^1\mathbf{Z}_i;\ \bigcup_{i=1}^{p_2} {}^2\mathbf{Z}_i\right],$$

or briefly $[w_1; w_2]$. We agree further that $[\ \ ; w_2]$ denotes the dichotomy defined by placing all p points into category w_2 and letting no point for category w_1. Out of all 2^p dichotomies, only a smaller number, as usual, may be realized by a hypersurface of a certain type, e.g. by an n-dimensional hyperplane.

To illustrate the concept of dichotomies by a very simple example, we give an account of all dichotomies of four points $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ lying in a plane (Fig. 1). There are altogether 16 dichotomies, namely $[\ \ ; \mathbf{A} \cup \mathbf{B} \cup \mathbf{C} \cup \mathbf{D}]$, $[\mathbf{A}; \mathbf{B} \cup \mathbf{C} \cup \mathbf{D}]$, … …, $[\mathbf{A} \cup \mathbf{B} \cup \mathbf{C} \cup \mathbf{D};\ \ ]$. Out of all these dichotomies, there are 14 dichotomies separable by a straight line, two dichotomies, namely $[\mathbf{A} \cup \mathbf{C}; \mathbf{B} \cup \mathbf{D}]$ and $[\mathbf{B} \cup \mathbf{D}; \mathbf{A} \cup \mathbf{C}]$, being unseparable by a curve of that type.

Definition 3. A set of p vectors in n-dimensional space is said to be in general position if every subset of n or fewer vectors is linearly independent.

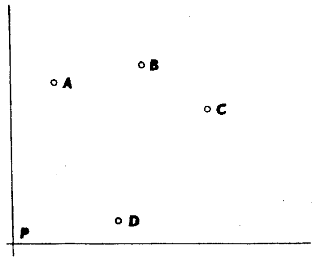

Fig. 1. Four points in general position in a plane.

Theorem 1. *Let us have in n-dimensional space two categories w_1 and w_2 defined by (4) and (5), let further be $p = p_1 + p_2$ and $\mathbf{Z}_{p+1}$ a point other than the origin of the space. Then the dichotomies*

$$[w_1 \cup \mathbf{Z}_{p+1};\, w_2]$$

and

$$[w_1;\, w_2 \cup \mathbf{Z}_{p+1}]$$

are homogeneously linearly separable if and only if

$$[w_1;\, w_2]$$

is homogeneously linearly separable by an $(n-1)$-hyperplane passing through $\mathbf{Z}_{p+1}$.

We shall follow here main ideas of the proof presented in [4]. According to Definition 2, there exist vectors $\mathbf{g}_1$ and $\mathbf{g}_2$ satisfying the following inequalities:

$$\begin{aligned}
&\mathbf{g}_1 \,.\, {}^1\mathbf{Z}_i > 0\,, \quad i = 1, \ldots, p_1\,,\\
&\mathbf{g}_1 \,.\, {}^2\mathbf{Z}_i < 0\,, \quad i = 1, \ldots, p_2\,,\\
&\mathbf{g}_1 \,.\, \mathbf{Z}_{p+1} > 0\,,\\
&\mathbf{g}_2 \,.\, {}^1\mathbf{Z}_i > 0\,, \quad i = 1, \ldots, p_1\,,\\
&\mathbf{g}_2 \,.\, {}^2\mathbf{Z}_i < 0\,, \quad i = 1, \ldots, p_2\,,\\
&\mathbf{g}_2 \,.\, \mathbf{Z}_{p+1} < 0\,.
\end{aligned}$$

Therefore for vector

$$\mathbf{g}_3 = (-\mathbf{g}_2 \,.\, \mathbf{Z}_{p+1})\, \mathbf{g}_1 + (\mathbf{g}_1 \,.\, \mathbf{Z}_{p+1})\, \mathbf{g}_2$$

 the following inequalities hold:

$$\boldsymbol{g}_3 \cdot {}^1\boldsymbol{Z}_i > 0\,, \quad i = 1, \ldots, p_1\,,$$
$$\boldsymbol{g}_3 \cdot {}^2\boldsymbol{Z}_i < 0\,, \quad i = 1, \ldots, p_2\,.$$

Since further

$$\boldsymbol{g}_3 \cdot \boldsymbol{Z}_{p+1} = 0\,,$$

the hyperplane dividing the two categories passes through $\boldsymbol{Z}_{p+1}$. Conversely, according to Definition 2, there exists a vector $\boldsymbol{g}_3$ satisfying

$$\boldsymbol{g}_3 \cdot {}^1\boldsymbol{Z}_i \quad > 0\,, \quad i = 1, \ldots, p_1\,,$$
$$\boldsymbol{g}_3 \cdot {}^2\boldsymbol{Z}_i \quad < 0\,, \quad i = 1, \ldots, p_2\,,$$
$$\boldsymbol{g}_3 \cdot \boldsymbol{Z}_{p+1} = 0\,.$$

Choosing a positive real number ε

$$\varepsilon < \frac{\boldsymbol{g}_3 \cdot {}^1\boldsymbol{Z}_i}{\left|\boldsymbol{Z}_{p+1} \cdot {}^1\boldsymbol{Z}_i\right|}\,, \quad i = 1, \ldots, p_1\,,$$
$$\varepsilon < \frac{\boldsymbol{g}_3 \cdot {}^2\boldsymbol{Z}_i}{\left|\boldsymbol{Z}_{p+1} \cdot {}^2\boldsymbol{Z}_i\right|}\,, \quad i = 1, \ldots, p_2\,;$$

and defining

$$\boldsymbol{g}_1 = \boldsymbol{g}_3 + \varepsilon\boldsymbol{Z}_{p+1}\,,$$
$$\boldsymbol{g}_2 = \boldsymbol{g}_3 - \varepsilon\boldsymbol{Z}_{p+1}\,,$$

we have

$$\boldsymbol{g}_1 \cdot \boldsymbol{Z}_{p+1} = \left(\boldsymbol{g}_3 + \varepsilon\boldsymbol{Z}_{p+1}\right) \cdot \boldsymbol{Z}_{p+1} > 0\,.$$

Further,

$$\boldsymbol{g}_1 \cdot {}^1\boldsymbol{Z}_i = \boldsymbol{g}_3 \cdot {}^1\boldsymbol{Z}_i + \varepsilon{}^1\boldsymbol{Z}_i \cdot \boldsymbol{Z}_{p+1} > \varepsilon\left(\left|{}^1\boldsymbol{Z}_i \cdot \boldsymbol{Z}_{p+1}\right| - {}^1\boldsymbol{Z}_i \cdot \boldsymbol{Z}_{p+1}\right) \geqq 0\,,$$
$$i = 1, \ldots, p_1\,,$$

and

$$\boldsymbol{g}_1 \cdot {}^2\boldsymbol{Z}_i = \boldsymbol{g}_3 \cdot {}^2\boldsymbol{Z}_i + \varepsilon{}^2\boldsymbol{Z}_i \cdot \boldsymbol{Z}_{p+1} < \boldsymbol{g}_3 \cdot {}^2\boldsymbol{Z}_i\left(1 - \frac{{}^2\boldsymbol{Z}_i \cdot \boldsymbol{Z}_{p+1}}{\left|{}^2\boldsymbol{Z}_i \cdot \boldsymbol{Z}_{p+1}\right|}\right) \leqq 0\,,$$
$$i = 1, \ldots, p_2\,,$$

where we used inequality $|a| - a \geqq 0$. Since it is possible to show in the same manner that

$$\boldsymbol{g}_2 \cdot \boldsymbol{Z}_{p+1} > 0\,,$$
$$\boldsymbol{g}_2 \cdot {}^1\boldsymbol{Z}_i \quad > 0\,, \quad i = 1, \ldots, p_1\,,$$
$$\boldsymbol{g}_2 \cdot {}^2\boldsymbol{Z}_i \quad < 0\,, \quad i = 1, \ldots, p_2\,,$$

according to Definition 2 the theorem is proved.

Lemma 1. *For natural numbers n, p the following identity is valid*:

$$\sum_{i=0}^{n-1}\binom{p-1}{i}+\sum_{i=0}^{n}\binom{p-1}{i}=\sum_{i=0}^{n}\binom{p}{i}. \tag{9}$$

Proof. For $n = 1$ and any natural p Lemma 1 is valid, namely

$$\binom{p-1}{0}+\binom{p-1}{0}+\binom{p-1}{1}=\binom{p}{0}+\binom{p}{1}.$$

Assuming that Lemma 1 has been established for $n - 1$, we show that it is valid also for n. In order to do it, we shall write (9) in the following way:

$$\sum_{i=0}^{n-2}\binom{p-1}{i}+\binom{p-1}{n-1}+\sum_{i=0}^{n-1}\binom{p-1}{i}+\binom{p-1}{n}=\sum_{i=0}^{n-1}\binom{p}{i}+\binom{p}{n}.$$

Since according to our induction assumption

$$\sum_{i=0}^{n-2}\binom{p-1}{i}+\sum_{i=0}^{n-1}\binom{p-1}{i}=\sum_{i=0}^{n-1}\binom{p}{i},$$

it is enough to prove that

$$\binom{p}{n}-\binom{p-1}{n}=\binom{p-1}{n-1},$$

which may be clearly seen from the definition of combination. Q.E.D.

Theorem 2. *Let us have p points $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_p$ in general position in n-dimensional Euclidean space. There are then*

$$C_\eta(p, n) = 2\sum_{i=0}^{n-\eta-1}\binom{p-1}{i}$$

dichotomies by hyperplanes passing through the origin of the space and η $(0 \leqq \eta \leqq n-1)$ points different from the origin of the space and such that all $\eta + p$ points are in general position.

This theorem as well as Theorem 3 may be found e.g. in [4]; we shall prove these theorems in a different form.

I. It may be readily verified that for $p = 1$ and for any natural n and $0 \leqq \eta \leqq$ $\leqq n - 1$ Theorem 2 is valid, namely

$$C_\eta(1, n) = 2\sum_{i=0}^{n-\eta-1}\binom{0}{i} = 2\,.$$

The theorem is also valid for $\eta = n - 1$ and for any natural p and n, namely

$$C_{n-1}(p, n) = 2\binom{p-1}{0} = 2\,.$$

II. Assuming that the theorem has been established for $p - 1$ points, we show that it is valid for p points, for any natural n and $0 \leqq \eta < n - 1$. Let us add to the $p - 1$ points a point $\boldsymbol{Z}_p$, which is in general position with all other points. All of the $C_\eta(p - 1, n)$ hyperplanes existing according to induction assumption may be divided into groups: those which pass through the point $\boldsymbol{Z}_p$ and those which do not. According to induction assumption there are $C_{\eta+1}(p - 1, n)$ hyperplanes passing through the origin of the space and the point $\boldsymbol{Z}_p$ and dividing the $p - 1$ points into two categories w_1 and w_2. According to Theorem 1 there exist for any hyperplane passing through $\boldsymbol{Z}_p$ two hyperplanes dividing

$$[w_1 \cup \boldsymbol{Z}_p;\, w_2]$$

and

$$[w_1;\, w_2 \cup \boldsymbol{Z}_p]\,,$$

respectively. Thus

$$C_\eta(p, n) = C_\eta(p - 1, n) + C_{\eta+1}(p - 1, n)\,.$$

Further, according to induction assumption

$$C_\eta(p, n) = 2\sum_{i=0}^{n-\eta-1}\binom{p-2}{i} + \sum_{i=0}^{n-\eta-2}\binom{p-2}{i},$$

and according to Lemma 1

$$C_\eta(p, n) = 2\sum_{i=0}^{n-\eta-1}\binom{p-1}{i}.$$

Q.E.D.

Putting $\eta = 0$ the following Theorem 3 immediately follows.

Theorem 3. *There are $C(p, n)$ homogeneously linearly separable dichotomies of p points in general position in n-dimensional Euclidean space, where*

$$C(p, n) = 2\sum_{i=0}^{n-1}\binom{p-1}{i}.$$

Theorem 4. *The number of homogeneously linearly separable dichotomies of p points in any position in n-dimensional space $C'(p, n)$ is always less than or equal to $C(p, n)$.*

Let us assume that there are p' points ($p' \leqq p$) in general position. If $p' = p$, Theorem 4 follows directly from Theorem 3. If $p' < p$, we may proceed as follows.

According to Theorem 3 there are $C(p', n)$ dichotomies for p' points. For any two categories w_1 and w_2 of the points, for which dichotomy exists, there exists also at least one of dichotomies

$$[w_1 \cup \mathbf{Z}_{p*}; w_2] \quad [w_1; w_2 \cup \mathbf{Z}_{p*}],$$

where $\mathbf{Z}_{p*}$ is any of remaining points $\mathbf{Z}_{p'+1}, \ldots, \mathbf{Z}_p$. If the point is not in general position, there exists a straight line defined by the origin of the space and one of the points $\mathbf{Z}_1, \ldots, \mathbf{Z}_{p'}$ such that $\mathbf{Z}_{p*}$ lies on it. Therefore there is no hyperplane passing through point $\mathbf{Z}_{p*}$ and dividing points $\mathbf{Z}_1, \ldots, \mathbf{Z}_{p'}$. Hence, according to Theorem 1, there exists at most one of the dichotomies

$$[w_1 \cup \mathbf{Z}_{p*}; w_2] \quad \text{and} \quad [w_1; w_2 \cup \mathbf{Z}_{p*}].$$

Thus the number of dichotomies remains the same and we may omit point $\mathbf{Z}_{p*}$ without changing the result. Therefore, the number of homogeneously linearly separable dichotomies is

$$C'(p, n) = C(p', n) \leqq C(p, n).$$

Note. The comparing of Theorem 2 and Theorem 3 yields the following identity valid for all natural p, n and $0 \leqq \eta < n - 1$:

$$C_\eta(p, n) = C(p, n - \eta).$$

Lemma 2. *Let us have p points $\mathbf{Z}_1, \ldots, \mathbf{Z}_p$ in n-dimensional space. There are then $D^s_\eta(p, n)$ dichotomies by s different hyperplanes passing through the origin of the space with the additional condition that one of the hyperplanes passes through $0 \leqq \eta \leqq n - 1$ points different from points $\mathbf{Z}_1, \ldots, \mathbf{Z}_p$ and provided that all $\eta + p$ points are in general position. $D^s_\eta(p, n)$ is defined as follows:*

$$D^s_\eta(p, n) = 2 \sum_{i=0}^{s(n-1)-\eta} \binom{p-1}{i}.$$

The Lemma will be established by the following inductive proof.

I. For $p = 1$ and for any natural s, n and $0 \leqq \eta \leqq n - 1$ the Lemma is valid, namely

$$D^s_\eta(1, n) = 2 \sum_{i=0}^{s(n-1)-\eta} \binom{0}{i} = 2.$$

II. Assuming that the Lemma has been established for $p - 1$ points $\mathbf{Z}_1, \ldots, \mathbf{Z}_{p-1}$ and for any natural s, n and $0 \leqq \eta \leqq n - 1$, we shall show that it is valid also for p points, any natural s, n and $0 \leqq \eta < n - 1$. Let us add to the $p - 1$ points a point $\mathbf{Z}_p$, which is in general position with all other points. All of the sets of hyperplanes, dividing points into two categories w_1 and w_2, may be divided into two groups: those which pass through the point $\boldsymbol{Z}_p$ and those which do not. According to Theorem 1, for any hyperplane passing through $\boldsymbol{Z}_p$ there exist two hyperplanes dividing

$$[w_1 \cup \boldsymbol{Z}_p;\, w_2]$$

and

$$[w_1;\, w_2 \cup \boldsymbol{Z}_p]\,,$$

respectively. Therefore also for any set of hyperplanes, one of which passes through the point $\boldsymbol{Z}_p$, there exist two sets of hyperplanes dividing

$$[w_1 \cup \boldsymbol{Z}_p;\, w_2]$$

and

$$[w_1;\, w_2 \cup \boldsymbol{Z}_p]\,.$$

Hence,

$$D_\eta^s(p, n) = D_\eta^s(p - 1, n) + D_{\eta+1}^s(p - 1, n)\,.$$

Using the induction assumption we get

$$D_\eta^s(p, n) = 2\left[\sum_{i=0}^{s(n-1)-\eta} \binom{p-2}{i} + \sum_{i=0}^{s(n-1)-\eta-1} \binom{p-2}{i}\right],$$

which yields according to Lemma 1

$$D_\eta^s(p, n) = 2\sum_{i=0}^{s(n-1)-\eta} \binom{p-1}{i}\,.$$

Since $n - 1$ points together with the origin of the space define just one hyperplane in n-dimensional space, the following identity must be valid for all natural p, n and s:

$$D_{n-1}^s(p, n) = D_0^{s-1}(p, n)\,.$$

Therefore, by establishing the Lemma for p points, any natural n, s and $0 \leqq \eta < n - 1$, we have also established it for $\eta = n - 1$. Thus the assumption of validity for $p - 1$ points leads to the validity of the Lemma for p points. Q.E.D.

Theorem 5. *Let us have p points in general position in n-dimensional space. There are then $D^s(p, n)$ dichotomies by s hyperplanes passing through the origin of the space, where*

$$D^s(p, n) = C(p, s(n - 1) + 1)\,.$$

Lemma 2 for $\eta = 0$ yields

$$D^s(p, n) = 2\sum_{i=0}^{s(n-1)} \binom{p-1}{i}\,.$$

Hence, according to Theorem 2

$$D^s(p, n) = C(p, s(n - 1) + 1) .$$

Q.E.D.

Theorem 6. *Let us have p points in any position in n-dimensional space. There are then at most $D^s(p, n)$ dichotomies by s hyperplanes passing through the origin of the space.*

Let us assume that there are p' points in general position. If $p' = p$, Theorem 6 follows directly from Theorem 5. If $p' < p$, we may proceed similarly as we did in the proof of Theorem 4. We should get

$$D'^s(p, n) = D^s(p', n) \leqq D^s(p, n) ,$$

where D' denotes the number of dichotomies of p points in any position. Q.E.D.

Definition 4. We shall call Φ-function any single-valued real function

$$\Phi(\mathbf{X}) = \sum_{r_1=0}^{n} \sum_{r_2=r_1}^{n} \cdots \sum_{r_t=r_{t-1}}^{n} g_{r_1 \ldots r_t} x_{r_1} \ldots x_{r_t} , \tag{10}$$

where

$$\mathbf{X} \equiv (x_1, \ldots, x_n)$$

and

$$x_0 = 1$$

for any natural n, t and real constants $g_{r_1 \ldots r_t}$ satisfying the following condition:

$$\left| \sum_{r_1=0}^{n} \sum_{r_2=r_1}^{n} \cdots \sum_{r_t=r_{t-1}}^{n} g_{r_1 \ldots r_t} \right| > 0 .$$

Note. The choice $t = 1$ leads to linear function

$$\Phi(\mathbf{X}) = \sum_{r=1}^{n} g_r x_r + g_0 ;$$

and taking additional condition $g_0 = 0$, we get homogeneously linear function

$$\Phi(\mathbf{X}) = \sum_{r=1}^{n} g_r x_r .$$

Both (11) and (12) define then a hyperplane in n-dimensional space, while the hyperplane defined by (12) passes through the origin of the space.

$$\sum_{r=1}^{n} g_r x_r + g_0 = 0 , \tag{11}$$

$$\sum_{r=1}^{n} g_r x_r = 0 . \tag{12}$$

Choosing $t = 2$, we get

$$\sum_{r_1=1}^{n} g_{r_1 r_1} x_{r_1}^2 + \sum_{r_1=1}^{n} \sum_{r_2=r_1-1}^{n} g_{r_1 r_2} x_{r_1} x_{r_2} + \sum_{r_1=1}^{n} g_{r_1 0} x_{r_1} + g_{00} = 0 ,$$

which is the equation of a hyperquadric in n-dimensional space.

Note. There are $\binom{n+t}{t}$ *constants* $g_{r_1 \ldots r_t}$ *in* Φ*-function defined by* (10).

Theorem 7. *There are at most* $C(p, m)$ *dichotomies of* p *points in* n*-dimensional space accomplished by surfaces*

$$\Phi(\boldsymbol{X}) = \sum_{r_1=0}^{n} \cdots \sum_{r_t=r_{t-1}}^{n} g_{r_1 \ldots r_t} x_{r_1} \ldots x_{r_t} = 0$$

(*i.e.* Φ*-dichotomies*), *where*

$$m = \binom{n+t}{t}.$$

To each point $\boldsymbol{Z}$ of the given n-dimensional Euclidean space there exists a point $\Phi(\boldsymbol{Z})$ in $\binom{n+t}{t}$-dimensional Φ-space. Therefore, to each set of points $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_p$ there exists a set of points $\Phi(\boldsymbol{Z}_1), \ldots, \Phi(\boldsymbol{Z}_p)$ in Φ-space; and for any homogeneously linearly separable dichotomy in Φ-space there exists a corresponding Φ-dichotomy in the given n-dimensional space. Since the number of homogeneously linearly separable dichotomies of p points in m-dimensional space is according to Theorem 3 at most $C(p, m)$, Theorem 7 is thus established.

Note. For $t = 1$

$$\Phi(\boldsymbol{X}) = 0 \tag{13}$$

defines a hyperplane in n-dimensional space not necessarily passing through the origin of the space. Thus the number of dichotomies by hyperplanes defined by (13) is equal to $C(p, n + 1)$.

Theorem 8. *There are at most* $C(p, s(m - 1) + 1)$ *dichotomies of* p *points in* n*-dimensional space accomplished by* s *surfaces*

$$\Phi^{(\beta)}(\boldsymbol{X}) = \sum_{r_1=0}^{n} \cdots \sum_{r_t=r_{t-1}}^{n} g^{(\beta)}_{r_1 \ldots r_t} x_{r_1} \ldots x_{r_t} = 0 , \quad \beta = 1, \ldots, s ,$$

where $m = \binom{n+t}{t}$.

For any dichotomy by s hyperplanes passing through the origin of the Φ-space there exists a dichotomy by s Φ-hypersurfaces in n-dimensional Euclidean space. Since according to Theorem 6 there are at most $C(p, s(m-1)+1)$ dichotomies in m-dimensional space, the Theorem is established.

In order to illustrate the proved theorems, a simple example will be presented. Let 4 points $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$ be given in a plane (Fig. 1). Out of all 16 dichotomies, only 8 dichotomies, namely $[\ ; \boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{C} \cup \boldsymbol{D}]$, $[\boldsymbol{A}; \boldsymbol{B} \cup \boldsymbol{C} \cup \boldsymbol{D}]$, $[\boldsymbol{A} \cup \boldsymbol{B}; \boldsymbol{C} \cup \boldsymbol{D}]$, $[\boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{C}; \boldsymbol{D}]$, $[\boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{C} \cup \boldsymbol{D};\ ]$, $[\boldsymbol{B} \cup \boldsymbol{C} \cup \boldsymbol{D}; \boldsymbol{A}]$, $[\boldsymbol{C} \cup \boldsymbol{D}; \boldsymbol{A} \cup \boldsymbol{B}]$, $[\boldsymbol{D}; \boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{C}]$, are separable by a straight line passing through $\boldsymbol{P}$, while using two straight lines, both passing through $\boldsymbol{P}$, another 6 dichotomies, namely $[\boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{D}; \boldsymbol{C}]$, $[\boldsymbol{A} \cup \boldsymbol{D}; \boldsymbol{B} \cup \boldsymbol{C}]$, $[\boldsymbol{A} \cup \boldsymbol{C} \cup \boldsymbol{D}; \boldsymbol{B}]$, $[\boldsymbol{C}; \boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{D}]$, $[\boldsymbol{B} \cup \boldsymbol{C}; \boldsymbol{A} \cup \boldsymbol{D}]$, $[\boldsymbol{B}; \boldsymbol{A} \cup \boldsymbol{C} \cup \boldsymbol{D}]$, become separable, and finally, using three straight lines, all of them passing through $\boldsymbol{P}$, there may be all 16 dichotomies separated.

Using a straight line without any restriction, there are, as has been shown above, 14 dichotomies separable by a curve of that type. Finally, using a quadratic curve without any restriction, all 16 dichotomies may be separated.

We shall now investigate a few typical pattern recognition systems with discriminant functions from the point of view of their separating capacity.

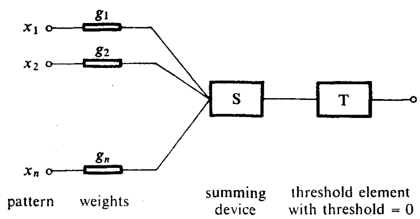

Fig. 2. Simple perceptron.

Example 1. Simple perceptron (Fig. 2) [5] uses a hyperplane defined by (12) and passing through the origin of the space. Therefore it follows directly from Theorem 4 that the number of separable dichotomies is at most $C(p, n)$.

Example 2. ADALINE (Fig. 3) [6] uses a hyperplane without any restriction, defined by (11). Therefore, as has already been shown, the number of separable dichotomies is at most $C(p, n + 1)$.

Example 3. The pattern recognition system with a quadric processor (Fig. 4) uses hyperquadrics as decision boundaries. Applying Theorem 7 and noting that

$$m = \tfrac{1}{2}(n-1)(n-2),$$

we get the maximum number of separable dichotomies equal to $C(p, \frac{1}{2}(n-1)(n-2))$.

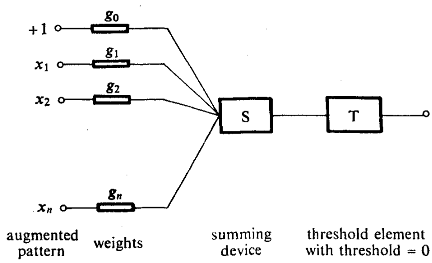

Fig. 3. ADALINE.

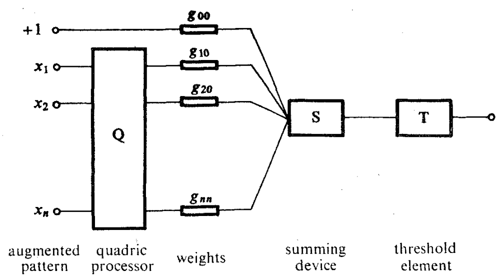

Fig. 4. Pattern recognition system with quadric processor.

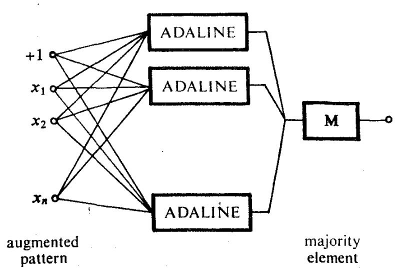

Fig. 5. Committee machine.

Example 4. The decision boundary of the committee machine shown in Fig. 5 is a set of s hyperplanes. If we admit any logic function of the element M, then according to Theorem 8 and taking into consideration that $m = n - 1$, we get immediately the maximum number of separable dichotomies of p points equal to $C(p, sn + 1)$. Since the element M realizes majority operation, only a few of all possible 2^s different logic functions are allowed, and therefore only a fraction of the maximum number of separable dichotomies may be realized by the system shown in Fig. 5. $C(p, sn + 1)$ serves thus as a rough estimation of the upper bound of the number of separable dichotomies.

CONCLUSION

We have studied some general aspects of a certain class of pattern recognition learning systems. The maximum number of separable dichotomies derived above is the main limitation of those systems. This limitation becomes very important in case of many sample patterns. Therefore many pattern recognition learning systems succesfully tested with a few sample patterns compared with the number of variable parameters may be expected to fail in case of requiring many sample patterns to be taken into consideration.

(Received November 11th, 1968.)

REFERENCES

[1] F. Rosenblatt: Principles of Neurodynamics. Spartan Books, 1962.

[2] R. C. Singleton: A Test for Linear Separability as Applied to Self-Organizing Machines. In: Self-Organizing Systems — 1962 (Yovits, Jacobi and Goldstein, Eds.), Spartan Books, 1962, 503—524.

[3] M. A. Айзерман, Э. М. Браверман, Л. Н. Розоноэр: Теоретические основы метода потенциальных функций в задаче об обучении автоматов распознаванию классов. Автоматика и телемеханика *25*, (Июнь 1964), 917—938.

[4] T. M. Cover: Geometrical and Statistical Properties of Systems of Linear Inequalities with Application in Pattern Recognition. IEEE Transactions on Electronic Computers *EC-14* (June 1965), 326—334.

[5] F. Rosenblat: A Comparison of Several Perceptron Models. In: Self-Organizing Systems — 1962 (Yovits, Jacobi and Goldstein, Eds.), Spartan Books, 1962, 463—484.

[6] B. Widrow, F. W. Smith: Pattern Recognizing Control Systems. In: Computer and Information Sciences (Tou and Wilcox, Eds.), Spartan Books, 1964, 288—317.

O separabilitě úloh při použití některých učících se systémů pro rozpoznávání obrazců

LUBOMÍR OHERA

Mnohé učící se systémy pro rozpoznávání obrazců jsou založeny na následujícím principu: Systém se učí na základě znalosti vzorových obrazců tak, že po skončení trenovací doby rozpoznává správně všechny vzorové obrazce. Příklady takových systémů jsou různé typy perceptronů, ADALINE, MADALINE a mnohé další. V článku jsou zkoumány systémy tohoto typu z hlediska schopnosti učit se různým situacím a jsou odvozeny odhady počtu situací, které mohou být systémem naučeny.

Ing. Lubomír Ohera, Fakulta jaderná a fyzikálně inženýrská ČVUT, Břehová 7, Praha 1.